Whole-genome sequencing and genomic-based acid tolerance mechanisms of Lactobacillus delbrueckii subsp. bulgaricus LJJ

xiaoyang pang (pangxiaoyang@163.com)
Chinese Academy of Agricultural Sciences https://orcid.org/0000-0003-1813-2039
Weixun Li
Chinese Academy of Agricultural Sciences Institute of Food Science and Technology
Lan Yang
Chinese Academy of Agricultural Sciences Institute of Food Science and Technology
Cheng Hu
Northeast Agricultural University, College of Food Science
Jing Lu
Chinese Academy of Agricultural Sciences Institute of Food Science and Technology
Shuwen Zhang
Chinese Academy of Agricultural Sciences Institute of Food Science and Technology
Obaroakpo Joy Ujiroghene
Chinese Academy of Agricultural Sciences Institute of Food Science and Technology
Changlu Ma
Chinese Academy of Agricultural Sciences Institute of Food Science and Technology
Cai Zhang
Henan University of Science and Technology, Laboratory of Environmene and Livestock Products
Jiaping Lv
Chinese Academy of Agricultural Sciences Institute of Food Science and Technology

Research article

Keywords: Lactobacillus delbrueckii subsp. Bulgaricus, acid-tolerant mechanism, comparative genomics, dapA, dapH, LysC

Posted Date: October 30th, 2019

DOI: https://doi.org/10.21203/rs.2.16590/v1

Abstract

Background: The probiotic efficacy and fermentative ability of Lactobacillus delbrueckii subsp. Bulgaricus (L. d. bulgaricus), a widely used probiotic, is majorly affected by its acid tolerance. In this study, a genome-wide sequence of a highly acid-tolerant L. d. bulgaricus LJJ was supposed to be determined, and we expect to find out the acid tolerance mechanism of L. d. bulgaricus LJJ by comparative genomics.

Results: Functional annotation and pathways of differential genes were determined using bioinformatics. The results in our study showed that the three genes dapA , dapH and lysC identified are implicated in the high acid tolerance of LJJ strain. Thus, they are potentially important as acid-tolerant genes of LJJ strain.

Conclusions: This study successfully revealed the acid tolerance mechanism of LJJ. Based on the previous research of LJJ in our laboratory, the successful analysis of the acid-tolerant mechanism of L. d. bulgaricus will further lay the foundation for the subsequent breeding of high acid-tolerant strains and greatly enhance their probiotic functions.

Keywords: Lactobacillus delbrueckii subsp. Bulgaricus ; acid-tolerant mechanism; comparative genomics; dapA ; dapH ; LysC

Background

As a recognized safe edible microorganism, *L. d. bulgaricus* is widely used in the industrial production of fermented dairy products such as yogurt and cheese. It functions in regulating the gut microflora of the human body, maintaining the intestinal micro-ecological balance, inhibiting the growth of pathogenic bacteria, and improving the immunity of the human body [1]. High acid tolerance *L. d. bulgaricus* can tolerate the low pH environment of the gastrointestinal tract, reach the small intestine in vivo, proliferate and exert its probiotic function in the body [2].

Therefore, the acid tolerance of *L. d. bulgaricus* determines whether it can play a probiotic role in the host. The current selection of acid-tolerant strains of *L. d. bulgaricus* and the analysis of acid-tolerant mechanisms have become hotspots in the study of lactic acid bacteria [3].

Many researchers have conducted several in-depth studies on the acid tolerance mechanism of lactic acid bacteria through multiple omics methods. The known acid-tolerant mechanisms of lactic acid bacteria include protein protection and repair [4], changes in cellular metabolic pathways [5], changes in cell membrane structural proteins [6], production of alkaline substances [7], Proton Motice Force theory [8], Two-component regulatory system theory, etc. [9]. However, studies on the acid-tolerant mechanisms of *L. d. bulgaricus* are relatively rare. The high acid tolerance strains of *L. d. bulgaricus* LJJ used in this research were isolated from a traditional fermented dairy product from China’s pastoral area. This study is based on genome-wide sequencing and comparative genomics analysis of the differences between the

L. d. bulgaricus LJJ and *L. d. bulgaricus* type strain ATCC11842 genome, and then analyze the unique genes to screen acid-tolerant genes initially. In seven strains of *L. d. bulgaricus* with different acid tolerance, statistical methods were used to verify the acid-tolerant genes initially screened, and then the acid-tolerant genes were identified. Comparison of acid tolerance of seven strains of *L. d. bulgaricus* is available in Additional file 1 [see Additional file 1]. Successfully analyzing the acid-tolerant mechanism of *L. d. bulgaricus* provides a theoretical basis for the subsequent selection of strains with high acid tolerance for improved probiotic functions.

Results

Functional genomic analysis of *L. d. bulgaricus* LJJ

The raw sequencing data of *L. d. bulgaricus* LJJ was subjected to quality control based on the three generations of the sequencing data. An additional file shows this in more detail [see Additional file 2]. The final result of genome assembly is available(Table 1). The gene prediction results of *L. d. bulgaricus* LJJ showed that a total of 2003 ORFs were predicted, with a total length of 1,598,697 bp, an average length of 798.15 bp, accounting for 84.54% of the genome [see Additional file 3,4]. The *L. d. bulgaricus* LJJ genome-wide circle map is shown in Figure 1. The results of GO annotation, COG classification, and KEGG classification are available in Additional files [see Additional file 5].

Comparative analysis of genomic sequences and basic features of *L. d. bulgaricus* LJJ and ATCC11842

Although the basic characteristics of the genome of strains of *L. d. bulgaricus* are similar, slight differences may also be possible among variations. These differences may be attributed to evolutionary variability as a result of environmental differences(Table 2).

Comparative analysis of homology between *L. d. bulgaricus* LJJ and ATCC11842 genes

The results of genome-wide homology analysis of *L. d. bulgaricus* LJJ and *L. d. bulgaricus* ATCC11842 are shown in Figure 2. There were a total of 1441 genes with homology, with 445 genes unique to *L. d. bulgaricus* LJJ, while 126 genes were unique to *L. d. bulgaricus* ATCC11842. The homology of *L. d. bulgaricus* LJJ and *L. d. bulgaricus* ATCC11842 gene is relatively high, reaching 84%. Also, a small number of genes having significant differences and no homology was found.

Collinear analysis of *L. d. bulgaricus* LJJ and ATCC11842 genomes

Genome-wide analysis of *L. d. bulgaricus* LJJ and type strain ATCC11842 is shown in Figure 3.

Our results showed that there was no large sequence rearrangement between the genomes of LJJ and ATCC11842, indicating that the collinearity was good. Insertion, deletion, inversion, and translocation between LJJ and ATCC11842 genome occur in three LCBs(Locally collinear blocks, LCB_S). Short gene fragments insertion or inversion implies that the two strains of *L. d. bulgaricus* may have undergone genetic recombination or metastasis during the evolution process. Therefore, it is speculated that the insertion or deletion of these fragments is likely to result in different acid tolerance between the two strains [10]. A follow-up investigation such as knocking out target genes is required to validate this postulation.

Comparative Gene Ontology(GO) annotations of unique genes in *L. d. bulgaricus* LJJ and ATCC11842

The gene ontology annotation of the unique genes of LJJ and ATCC11842 is mainly focused on three levels: cellular component, molecular function, and biological process. In comparison to ATCC11842, LJJ has an advantage in the classification of cellular components, such as cell parts and membrane parts [see Additional file 6]. In the classification of molecular function, sub-classification, molecular transcription activity, binding, catalytic activity, and transporter activity plays important roles. For the classification of biological processes, its regulation, cellular processes, response stimuli, single-organism processes, and metabolic processes are necessary.

In these aspects, LJJ is advantageous probably because they have increased H^+-ATPase activity on the cell membrane during acidic environments, and also play an important role in the acid tolerance process [11].

The synthesis of cell membrane is an important cell biological property, in which a large number of genes related to the function of cell components are required to be expressed [12]. The acid tolerance is the most important biological function of lactic acid bacteria against acid stress. The acid stress process involves stimulating cell information interaction, gene expression regulation, substance transportation metabolism, etc. [13]. Consequently, biological processes are improved as a result of the relation between functional genes and molecular functions.

Cluster of Orthologus groups (COG) comparative analysis of the unique genes in *L. d. bulgaricus* LJJ and ATCC11842

In comparison with ATCC11842, LJJ has a larger proportion in the following aspects: (L) replication, recombination, and repair; (M) cell membrane biosynthesis and outer membrane proteins; (C) energy production and conversion; (E) amino acid transport and Metabolism; (V) defense mechanisms [see Additional file 6]. The number of functional genes related to cell protection in LJJ is significantly higher

than those of ATCC11842 and could be responsible for its higher acid tolerance level. Although LJJ naturally habits acid-stressed environments, it could, however, have difficulty in surviving normally during insufficient nutrients and energy [14]. In the absence of strong membrane and membrane protein synthesis or self-healing systems, it is difficult for even acid-tolerant probiotics to survive in an acidic environment [6]. GO annotations and COG classification Comparative analysis of unique genes in *L. d. bulgaricus* LJJ and ATCC11842.

Preliminary screening of *L. d. bulgaricus* LJJ acid-tolerant genes

Based on the results of the comparative genomics, the acid-tolerant gene was initially screened from the unique genes of LJJ and ATCC11842(Table 3). Most of the acid-tolerant genes initially screened are related to amino acid metabolism, suggesting the unique role of amino acids in bacterial acid tolerance.

PCR amplification and verification of acid-tolerant genes

Using the designed P1, P2….P16 as amplification primers, and genomes of 7 different *L. d. bulgaricus* strains as templates, 16 possible acid-tolerant genes of LJJ were amplified by PCR. The amplification results are shown in Figure 4.The result showed that the 7 different strains of *L. d. bulgaricus* contained the target genes corresponding to the P1, P7, P10, P11, P14, and P16 primers. Most fragments were also similar in size, implying that the target genes had no fragment insertion and loss. However, the target gene was not successfully amplified, indicating that the strain does not contain the gene fragment. A successful amplification would imply that the amplified fragments were different, thus proving that the genes had insertion and deletion in different strains.

Sequenced acid-tolerant genes and acid-tolerant-related gene verification in *L. d. bulgaricus*

The amplified fragment products were sequenced(Table 4). The sizes of the gene fragments were clearly expressed by the same gene in different strains, and are significantly different. Some fragments are obviously shorter than the length of the target gene fragment, which may have occurred as a result of gene deletion during evolution. However, a few other fragments were significantly longer than those of the target gene. The possible reason for these variations is that the fragment insertion occurs during the evolution of the gene, resulting in different fragment sizes, which ultimately leads to genetic functional changes. Consequently, the same gene is expressed differently in different strains, in such a way that some strains contain the gene, while others do not. Therefore, it is speculated that the related acid-tolerant protein cannot be normally expressed due to the deletion of the gene thus the acid-tolerant metabolic pathway functions abnormally resulting in a relatively weak acid tolerance.

Sequence alignment analysis of acid-tolerant genes

The sequence of selected acid-tolerant genes (No. 6, 12 and 13) were further analyzed, and the acid-tolerant genes *dapA, dapH* and *lysC* in different strains were also analyzed to identify the differences in their gene expressions. The results of the alignment analysis are shown in Figure 5a, 5b, and 5c

Figure 5(a) shows the alignment of *dapA* in different strains. The sequence alignment showed that large fragment deletions of *dapA* occur in JB, M13, ATCC11842, and GMC. It is speculated that the active center of the enzyme may be inactivated due to the deletion of the DNA fragment, resulting in the lysine metabolic pathway cannot be catalyzed to synthesize lysine. Thus, the acid tolerance of JB, M13, ATCC11842, and GMC is considered weak. Figure 5(b) and 5(c) show the amplification sequence alignment of *dapH* and *LysC* in different strains. After alignment, it was found that the expression results of the genes were consistent in the three strains LJJ, SY3, and YL5 with relatively strong acid tolerance. Based on the inferences from our study, *dapH* and *lysC* could be regarded as acid-tolerant genes of *L. d. bulgaricus* LJJ

Analysis of metabolic regulation of acid-tolerant genes of *L. d. bulgaricus* LJJ

According to previous research, the acid tolerance of lactic acid bacteria includes the regulation of intracellular H^+ [15], the regulation of cell membrane [16], stress response protein expression [17]. Likewise, the regulation of amino acid metabolism plays an important role in the regulation of intracellular pH, such as the glutamate decarboxylase system [18], Arginine deiminase(ADI) pathway [18].

In this study, we obtained three genes (*dapA, dapH,* and *lysC)* related to the acid tolerance of *L. d. bulgaricus* LJJ. The analysis of the acid-tolerant metabolic regulation was performed by KEGG. The regulation process of lysine synthesis by *dapA, dapH,* and *lysC* is shown in Figure 6.

Three important acid-tolerant genes *dapA, dapH,* and *lysC* mainly regulate the synthesis of lysine and also participate in the lysine metabolic pathway. The *lysC* gene encodes an aspartate kinase, while the *dapA* gene encodes a 4-hydroxy-tetrahydrodipicolinate synthase. These two enzymes are important for the synthesis of lysine from aspartate and are involved in all metabolic pathways such as succinyl-DAP (diaminopimelate) pathway, acetyl-DAP pathway, DAP dehydrogenase pathway, and DAP aminotransferase pathway. The *dapH* gene encodes a tetrahydrodipicolinate N-acetyltransferase, a key enzyme in the acetyl-DAP pathway.

Discussion

Analysis of KEGG metabolic pathways for *dapA, dapH,* and *lysC* indicated that they are all involved in lysine synthesis. Among them, *lysC* and *dapA* are the major key genes in the synthesis of lysine and are

involved in all metabolic pathways, such as succinyl-DAP pathway, acetyl-DAP pathway, DAP dehydrogenase pathway, and DAP aminotransferase pathway.

Moreover, *lysC* encodes aspartate kinase, which converts aspartate to phosphorylated aspartate. Phosphorylated aspartate is not only involved in lysine metabolism but also involved in amino acid metabolisms such as glycine, serine, threonine cysteine and phenylalanine, 2-oxocarboxylate Acid metabolism and various secondary metabolic pathways [19]. It is one of the important genes that enable strains to cope with changes in the external environment. This study suggests that the strong acid-tolerance ability in *L. d. bulgaricus* LJJ is strongly related to the three genes (*dapA, dapH,* and *lysC)* identified.

Although we have not found the gene *lysC* in ATCC11842, we found that its compensatory gene *ask,* which belongs to the same aspartic kinase family as *lysC.* The interaction about the compensation gene *ask* is available in Additional files [see Additional file 8]. Both *dapA* and *dapH* are unique genes, and no compensation has been found in ATCC11842, which may result in lysine synthesis always being at a lower level.

As a kind of basic amino acid, lysine can regulate the acid-base balance and help the strains' survival under acid stress. The acetylation modification of a protein generally refers to the transfer of an acetyl group from Acyl-CoA (acetyl coenzyme A) to a protein-specific lysine ε-amino group to form an acetylated lysine. The acetylation modification of lysine is also an important way to regulate the metabolic activities of prokaryotes [20]. Most of the proteins with acetylation are cytoplasmic proteins, which play a role closely related to cell metabolism, transcription, translation, etc. The acetylation of lysine can sense the energy state of cells and regulate the metabolism of cells through Acyl-CoA and NAD^+ (Nicotinamide adenine dinucleotide). It is a conservative regulation of metabolism in both eukaryotic and prokaryotic organisms [21]. Recent researches have shown that the addition of amino acids such as glutamic acid, arginine, lysine, and isoleucine can also increase the acid tolerance of cells [22].

Based on comparative genomics, we screened for acid-tolerant genes at the DNA level. However, the transcription level of an acid-tolerant gene is also an important factor affecting acid tolerance. Therefore, further analysis of acid tolerance depends on the transcriptomics data.

Conclusion

In this study, the acid tolerance of different strains of *L. d. bulgaricus* compared, and then screened to verify strains with strong acid tolerance (LJJ), and weak tolerance (ATCC11842). Genome-wide sequencing was performed on *L. d. bulgaricus* LJJ, to further predict the specific genes of the strain. 2003 ORFs with a total length of 1,598,697 bp and an average length of 798.15 bp, accounting for 84.54% of the genome was detected in our results. The GO gene function was annotated on the predicted genes, and the LJJ predicted gene was annotated with more genes in the binding, catalytic activity, cell part, cell, cellular process, biological adhesion, and metabolic process; which further confirmed that the

biological characteristics were strong in this aspect. Analysis of the COG functional classification of proteins predicted by the *L. d. bulgaricus* LJJ genome indicated that among the 2003 genes, 1251 genes were assigned functional annotations with significant biological significance. our results also showed that the genes were outstanding in amino acid transport metabolism, translation, ribosome structure and biosynthesis, recombination, repair, cell wall, and cell membrane synthesis. Based on the GO and COG annotation results, we further analyzed the classification of LJJ's metabolic pathways, which were enriched in the biosynthesis and transport of metabolites.

Through the comparative analysis of homology and collinearity, between genomes of *L. d. bulgaricus* LJJ(strong acid-tolerant) and *L. d. bulgaricus* ATCC11842(weak acid-tolerant), the related regulation pathway of acid-tolerant genes were explored. A comparative analysis of GO annotations and COG classifications between unique genes *of L. d. bulgaricus* LJJ and *L. d. bulgaricus* ATCC11842 was also determined, and the results showed that the number of genes in LJJ during acid metabolism, biofilm synthesis, and protein synthesis was significantly higher than those of ATCC11842. of the analysis A comprehensive analysis of the homologous clustering of LJJ and ATCC11842 strains revealed unique acid-tolerant genes from LJJ..Overall, sixteen genes associated with LJJ acid tolerance were initially screened.

Through the amplification and sequencing analysis of 16 possible acid-tolerant genes in 7 strains, the acid-tolerant genes were screened and identified as *dapA, dapH,* l*ysC.*In conclusion, *L. d. bulgaricus* LJJ may have a higher acid tolerance during an increase in lysine synthesis. Lysine is always at high levels of synthesis, thus aiding most lactic acid strains in coping with the acidity of external environment.

Methods

Source of bacteria

L. d. bulgaricus type strain ATCC11842 (NC_008054) was purchased from American Type Culture Collection, *L. d. bulgaricus* LJJ (CGMCC No.2687) was isolated from China General Microbiological Culture Collection Center, while *L. d. bulgaricus* (YL5, SY3, GMC, JB, M13) were obtained from some kinds of dairy products.

DNA sequencing

Whole-genome sequencing of *L. d. bulgaricus* LJJ was performed by PacBio sequencing and assembled by SOAPdenovo software based on sequencing data. The whole-genome sequence of *L. d. bulgaricus* ATCC11842 was obtained from Genbank (NC_008054).

L. d. bulgaricus LJJ gene function analysis and annotation

The Glimmer 3.02 software was used to predict the gene of the bacteria, and the ORF(open reading frame) with a length of more than 30 amino acids was selected. The predicted gene was searched for homology in the NCBI NR database, and the results were sorted to obtain the ORF of the LJJ genome[23].

Genome-wide circle map of *L.d. bulgaricus* LJJ

In this study, Circos V0.64 software was used in drawing the genome-wide circle.

Comparative genomics study of *L. d. bulgaricus* LJJ and type strain ATCC11842

The comparative genomic analysis involves the genome-wide comparison, genomic homology, and collinearity, as well as determining differential genes between *L.d bulgaricus LJJ* and *L. d. bulgaricus* type strain (ATCC11842). GO (Gene Ontology) annotation and COG (Cluster of Orthologus groups) functional cluster analysis were performed on unique genes, and candidate genes probably involved in acid-tolerant regulation were searched for from LJJ unique genes.

Verification of acid-tolerant genes

First, based on the results of the comparative genome, the acid-tolerant gene was initially screened from the unique genes of LJJ and ATCC11842. Then we designed primers for the initially screened acid-tolerant genes(Table 5). The whole genome of seven *L. d. bulgaricus* strains with different acid tolerance were used as templates to carry out gene amplification. According to the amplification results and the strength of the acid tolerance of the seven strains, the genes with slight differences were excluded. Strains with significant differences in genes were identified as acid-tolerant related genes.

If a gene is expressed in all strains and the size of the expressed gene fragments is different when it is matched with the difference in acid tolerance of the strain, then the gene can be preliminarily identified as a key gene for acid tolerance. If a gene is expressed only in a strain with strong acid tolerance and is not expressed in a strain with weak acid tolerance, the gene can also be preliminarily identified as a key gene for acid tolerance.

Amplification Sequence Analysis of Acid-tolerant Genes

The sequence of the selected acid-tolerant genes was analyzed by MEGA7, and the deletion or insertion was analyzed in order to identify the key reasons for the acid tolerance of the strains.

Analysis of KEGG(Kyoto Encyclopedia of Genes and Genomes) pathway of acid-tolerant gene

After validation of acid-tolerant genes, We searched the metabolic pathway of acid-tolerant genes in the KEGG database, revealing the reason why LJJ has strong acid tolerance at the genetic level

Abbreviations

L. d. bulgaricus: Lactobacillus delbrueckii subsp. Bulgaricus

GO: Gene Ontology *COG:* Cluster of Orthologus groups

CDS: Coding sequence *ORF:* Open reading frame

KEGG: Kyoto Encyclopedia of Genes and Genomes

$LCB_{S:}$ Locally collinear blocks,

ADI: Arginine deiminase *Acyl-CoA:* acetyl coenzyme A

NAD^{+}*:* Nicotinamide adenine dinucleotide

L-Asp: L-Aspartate *4-P-L-Asp:* 4-Phospho-L-aspartate

L-Asp−4-Sem: L-Aspartate 4-semialdehyde

HTPA: (2S,4S)−4-Hydroxy−2,3,4,5-tetrahydrodipicolinate

L−2,3,4,5-Tet: L−2,3,4,5-Tetrahydrodipicolinate

L−2-Ace−6-Oxo: L−2-Acetamido−6-oxoheptanedioate

N6-Ace-L−2,6-Dia: N6-Acetyl-L−2,6-diaminoheptanedioate

LL−2,6-Dia: LL−2,6-Diaminoheptanedioate

Meso-Dia: Meso-Diaminoheptanedioate

L-lys: L-Lysine

L−2-Ami−6-Oxo: L−2-Amino−6-oxoheptanedioate

N-Suc-L−2-Ami−6-Oxo: N-Succinyl-L−2-amino−6-oxoheptanedioate

N-Suc-L−2,6-Dia: N-Succinyl-L−2,6-diaminopimelate

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

Whole-genome sequence information of *Lactobacillus delbrueckii subsp. Bulgaricus* model strain ATCC11842 is available in GeneBank, (the accession number is NC_008054, https://www.ncbi.nlm.nih.gov/nuccore/NC_008054.1/). The datasets supporting the results of this article are included within the article and its additional files.

Competing interests

All authors have read and approved the manuscripts. The authors declare that they have no competing interests.

Funding

This work was financially supported by National Key R&D Program of China (2017YFC1600903), National Natural Science Foundation of China (Grant No.31871833), The Science and Technology Project of Beijing Education Committee (No. KM201812448003) and Beijing Municipal Education Commission Science & Technology Plan Projects (KM201812448003).

Acknowledgments

The authors would like to thank Na Chen for providing some pretty inspiration.

Authors' contributions

X.Y P, J L, S.W Z, and J L designed the study. W.X L, L Y, X.Y Pang and O.J U wrote the original draft. W.X L and L Y performed experiments. C.L M performed a part of bioinformatics analysis. All authors

contributed with writing, reviewing and editing. All authors read and approved the final version of the manuscript.

References

1.Knight DJW, Girling KJ: *Gut flora in health and disease. Lancet* 2003, *361(*9371):1831–1831.

2.Duwat P, Cesselin B, Sourice S, Gruss A: *Lactococcus lactis, a bacterial model for stress responses and survival. International Journal of Food Microbiology* 2000, *55(*1–3):83–86.

3.Cotter PD, Hill C: *Surviving the acid test: Responses of gram-positive bacteria to low pH. Microbiology and Molecular Biology Reviews* 2003, *67(*3):429-+.

4.Jan G, Leverrier P, Pichereau V, Boyaval P: *Changes in protein synthesis and morphology during acid adaptation of Propionibacterium freudenreichii. Applied and Environmental Microbiology* 2001, *67(*5):2029–2036.

5.De Angelis M, Gobbetti M: *Environmental stress responses in Lactobacillus: A review. Proteomics* 2004, *4(*1):106–122.

6.Wu C, Zhang J, Wang M, Du G, Chen J: *Lactobacillus casei combats acid stress by maintaining cell membrane functionality. Journal of Industrial Microbiology & Biotechnology* 2012, *39(7*):1031–1039.

7.Small PLC, Waterman SR: *Acid stress, anaerobiosis and gadCB lessons from Lactococcus lactis and Escherichia coli. Trends in Microbiology* 1998, *6(*6):214–216.

8.Koebmann BJ, Nilsson D, Kuipers OP, Jensen PR: *The membrane-bound H+-ATPase complex is essential for growth of Lactococcus lactis. Journal of Bacteriology* 2000, *182(*17):4738–4743.

9.Azcarate-Peril MA, McAuliffe O, Altermann E, Lick S, Russell WM, Klaenhammer TR: *Microarray analysis of a two-component regulatory system involved in acid resistance and proteolytic activity in Lactobacillus acidophilus. Applied and Environmental Microbiology* 2005, *71(*10):5794–5804.

10.Ming H, Xu D, Guo Z, Liu Y: *Adaptive Evolution of Lactobacillus casei under Acidic Conditions Enhances Multiple-stress Tolerance. Food Science and Technology Research* 2016, *22(*3):331–336.

11.Lorca GL, de Valdez GF: *Acid tolerance mediated by membrane ATPases in Lactobacillus acidophilus. Biotechnology Letters* 2001, *23(*10):777–780.

12.Anisimova EA, Yarullina DR, Ilinskaya ON: *Antagonistic activity of lactobacilli isolated from natural ecotopes. Microbiology* 2017, *86(*6):708–713.

13.Kandola S, Teotia UvS, Kumar R, Mishra AK, Singh A: *Investigation of acid tolerance attribute of various Lactobacillus casei group strains. Indian Journal of Animal Research* 2016, *50(*2):190–193.

14.Huang R, Pan M, Wan C, Shah NP, Tao X, Wei H: *Physiological and transcriptional responses and cross protection of Lactobacillus plantarum ZDY2013 under acid stress. Journal of Dairy Science* 2016, *99(*2):1002–1010.

15.Wu C, Zhang J, Du G, Chen J: *Aspartate protects Lactobacillus casei against acid stress. Applied Microbiology and Biotechnology* 2013, *97(*9):4083–4093.

16.Paradeshi JS, Patil SN, Koli SH, Chaudhari BL: *Effect of copper on probiotic properties of Lactobacillus helveticus CD6. International Journal of Dairy Technology* 2018, *71:*204–212.

17.Chen X, Sun Z, Meng H, Zhang H: *The acid tolerance association with expression of H+-ATPase in Lactobacillus casei. International Journal of Dairy Technology* 2009, *62(*2):272–276.

18.Schwenk WS, Donovan TM: *A Multispecies Framework for Landscape Conservation Planning. Conservation Biology* 2011, *25(*5):1010–1021.

19.Lee KH, Park JH, Kim TY, Kim HU, Lee SY: *Systems metabolic engineering of Escherichia coli for L-threonine production. Molecular Systems Biology* 2007, *3.*

20.Kouzarides T: *Acetylation: a regulatory modification to rival phosphorylation? Embo Journal* 2000, *19(*6):1176–1179.

21.Wang Q, Zhang Y, Yang C, Xiong H, Lin Y, Yao J, Li H, Xie L, Zhao W, Yao Y *et al: Acetylation of Metabolic Enzymes Coordinates Carbon Source Utilization and Metabolic Flux. Science* 2010, *327(*5968):1004–1007.

22.Reeve BWP, Reid SJ: *Glutamate and histidine improve both solvent yields and the acid tolerance response of Clostridium beijerinckii NCP260. Journal of Applied Microbiology* 2016, *120(*5):1271–1281.

23.Kelley DR, Liu B, Delcher AL, Pop M, Salzberg SL: *Gene prediction with Glimmer for metagenomic sequences augmented by classification and clustering. Nucleic Acids Research* 2012, *40(*1).

Tables

Table 1 Data statistics of third-generation sequencing

Sample	*L. d. bulgaricus* LJJ
Total reads num	103742
Total bases(bp)	582722054
Average length(bp)	5617
Number of all scaffolds	1
Length of genome(bp)	1891087
G+C content(%)	49.493

Table 2 Comparison of basic information of the *L. d. bulgaricus* genome

Strain	LJJ	ATCC11842
Accession number	——	NC_008054.1
Gene num	2003	1961
Gene total length(bp)	1598697	1864998
Gene average length(bp)	798.15	951.04
GCcontent(%)	51.26%	49.7%
Number of CDSs	1886	1836
Ribosome RNA	24	27
Number RNA	89	95
Plasmids	——	——
Genomic islands	18	——
repeat	49	41

“—” indicates no related information.

Table 3 Preliminary screening of acid-tolerant genes in *L. d. bulgaricus* LJJ

Number	Gene name	Pathway	Acid-tolerant mechanism theory
1	*TC.APA*	Catalyzing the formation of ammonia	Catalyzing the formation of ammonia
2	*argH*	Arginine metabolic pathway	Basic amino acid production
3	*asnB*	Alanine, aspartate and glutamate metabolism	Transamination
4	*metE*	Methionine anabolism	Alkali and biofilm synthesis mechanism
5	*patA*	Lysine metabolic pathway	Alkali synthesis mechanism
6	*dapA*	Lysine metabolic pathway	Alkali synthesis mechanism
7	*urdA*	Cycle of urea	Alkali synthesis mechanism
8	*mumM*	Cysteine, methionine metabolic pathway	Biofilm synthesis mechanism
9	*argC*	Two-component regulatory system	Biofilm synthesis mechanism
10	*LuxS*	Two-component regulatory system	Biofilm synthesis mechanism
11	Cation transporter	Maybe H^+ transporter	Transfer H+
12	*dapH*	Lysine metabolic pathway	Alkali synthesis mechanism
13	*lysC*	Aspartic acid, lysine metabolic pathway	Alkali synthesis mechanism
14	*proB*	Arginine, proline metabolic pathway	Alkali synthesis mechanism
15	*RecA*	——	Repair DNA damage
16	*arcD*	Arginine metabolic pathway	Alkali synthesis mechanism

“—” indicates no related information.

Table 4 Sequenced acid-tolerant genes

No.	Name/（bp）	LJJ[1]	SY3[2]	YL5[3]	GMC[4]	M13[5]	JB[6]	ATCC11842[7]
1	*TC.APA*	437	448	432	560	461	561	560
2	*argH*	1250	1221	__	911	904	1253	__
3	*asnB*	1082	1082	__	1060	1061	1073	1064
4	*metE*	2170	__	2149	2000	2145	2147	__
5	*patA*	1075	1049	1057	968	1058	__	__
6	*dapA*	793	792	692	485	498	483	479
7	*urdA*	557	563	562	560	554	547	550
8	*mumM*	564	551	539	__	551	218	326
9	*argC*	1190	1080	__	946	__	981	1082
10	*LuxS*	553	552	450	459	453	470	447
11	Cation transporter	500	486	__	466	486	448	__
12	*dapH*	624	622	714	__	__	__	__
13	*lysC*	1134	1127	1374	__	__	__	__
14	*proB*	391	381	472	404	391	378	380
15	*RecA*	884	883	521	885	884	885	883
16	*arcD*	468	468	496	492	495	491	506

“—” indicates that the strain does not contain the gene; “1” and “2” in the upper right corner of the strain indicate the order of acid tolerance. According to the screening principle of acid-tolerant genes, genes 6, 12 and 13 are directly related to the strength of acid tolerance.

Table 5: Amplification primers for acid-tolerant genes

Number	Primer name	Primer sequence(5'→3')	Annealing temperature/℃
P1	*TC.APA*	F: AAGTTTGGTGGCAGTTCCCT R: GCTGAATGGACGCGTTTTGT	59℃
P2	*argH*	F:TGACCTGCATAACTGCCTGG R: TTTCTTCCGCCAGCTTGTCT	58℃
P3	*asnB*	F:CCGCTTGGCTGCTTCTTTTT R:TCCTGCCTGGCCACTACTAT	57℃
P4	*MetE*	F:TCCAGTGCATCAATCGCCTT R:AACCGGCTTTGCGAGAAAAC	55℃
P5	*patA*	F:TAACCTAGCCGGTAGCCAGT R: TTGAAGCAGACGACTCCCAC	57℃
P6	*dapA*	F:AACTGCCGTAAAGTGAGCGA R:TGCAACGGCTTTGTGATTGG	56℃
P7	*urdA*	F:GGCGATTGAAAAGCAAGGCA R: GATACTGCCGTACCTGAGCC	59℃
P8	*mmuM*	F:TACTGCTGGAGCTGAACTGG R:AATTTGGCGCAATCTTGGCA	55℃
P9	*agrC*	F:TCATTGAGTTGGCCGCTTCT R: ACGGTCTGACCTGCTCCTAT	58℃
P10	*Luxs*	F:CGTCGAGTGACCCCTTCAAA R: GCCATTCCTACTGCAGGGTT	57℃
P11	Cation transporter	F: AGGCGATGATGGTGGTCAAG R: TTGCCCAGGACTTCCACAAG	57℃
P12	*dapH*	F:TGGTCACGGATAATAGCGCC R: CTTTCCCAGAAGGCACGGAT	58℃
P13	*lysC*	F:CAGCGCTACGTTGAAATCGG R:AGTCAGCGTGGAAACCTCTG	58℃
P14	*proB*	F:TGATCGAGGAATACACCCGC R: TTCTCCAGTTCAACGACGCC	58℃
P15	*RecA*	F:GAACTTCGGTAAAGGCGCGG R: CAGGAGCTGCCTTCTTAGCC	60℃
P16	*arcD*	F:TGCACAACCAGCTTTGGGTA R: TTCTGCCTTGTTCACGGACG	56℃

Figures

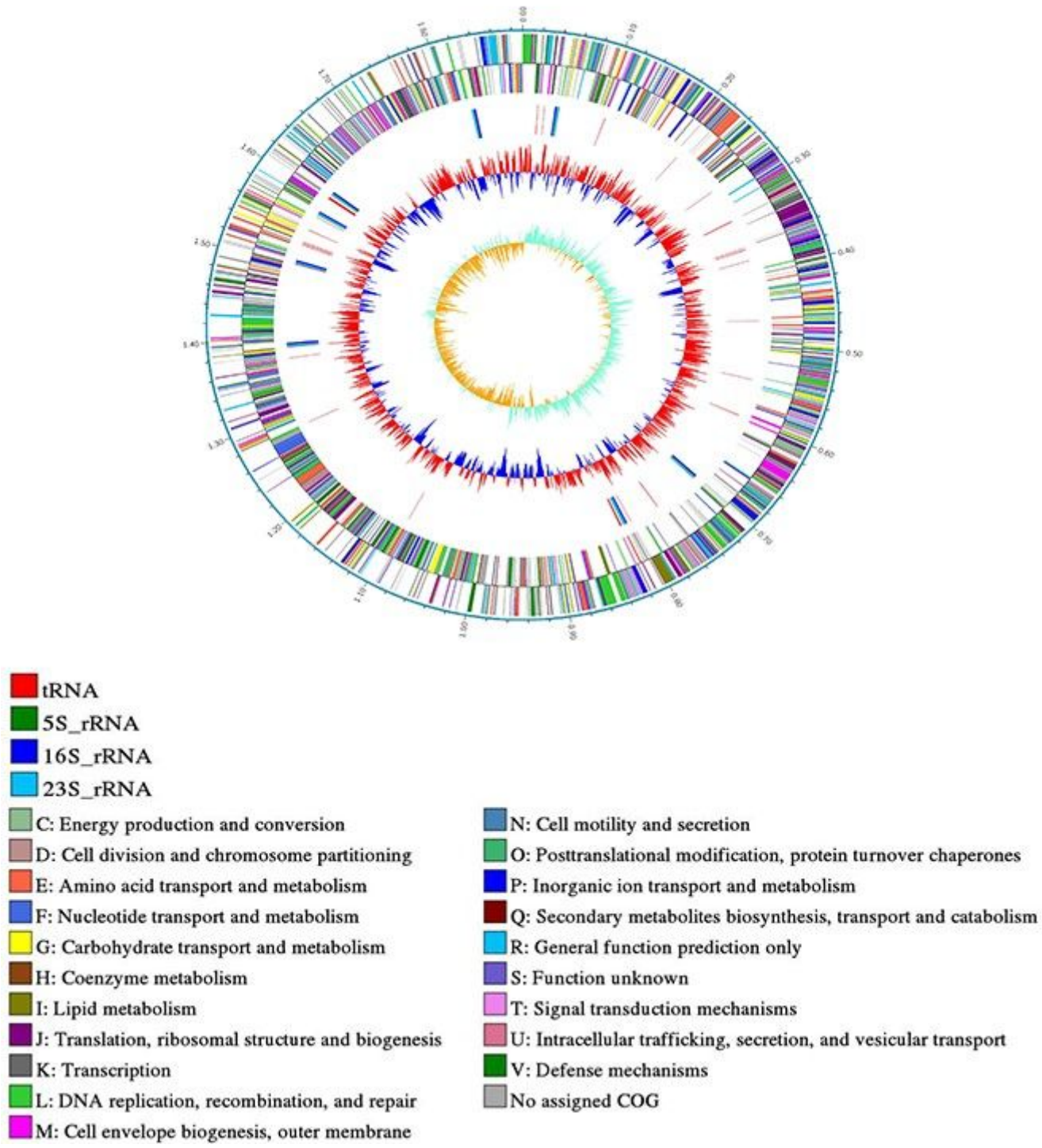


Figure 1

LJJ gene circle map The outer region of the gene-wide circle map represents the genome size; the second and third regions represent CDS (where different colors are different classes of COG); the fourth region represent the rRNA and tRNA; while G+C (%) is represented in the fifth region (where red indicates that the G+C content is higher than the average G+C in the whole genome, and blue indicates that the content of G+C is lower than the average G+C content of the whole genome). The innermost region is the GC skew value.

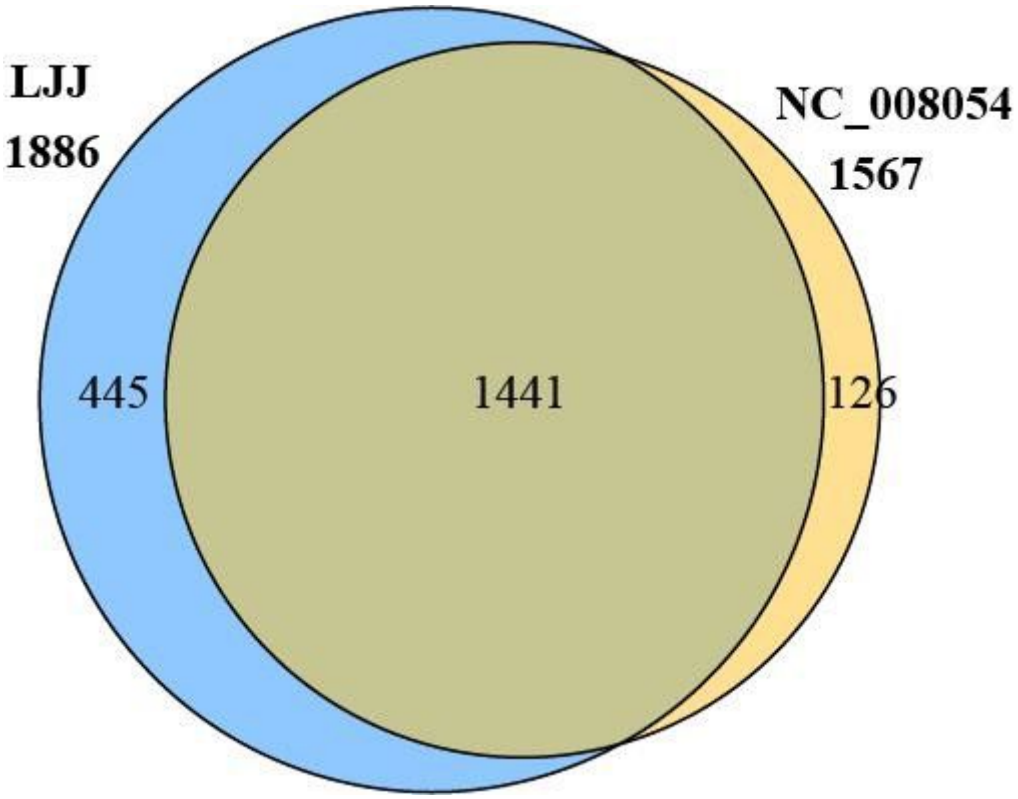


Figure 2

Homology comparism between LJJ and ATCC11842 There were a total of 1441 genes with homology, with 445 genes unique to L. d. bulgaricus LJJ, while 126 genes were unique to L. d. bulgaricus ATCC11842. The homology of L. d. bulgaricus LJJ and L. d. bulgaricus ATCC11842 gene is relatively high, reaching 84%.

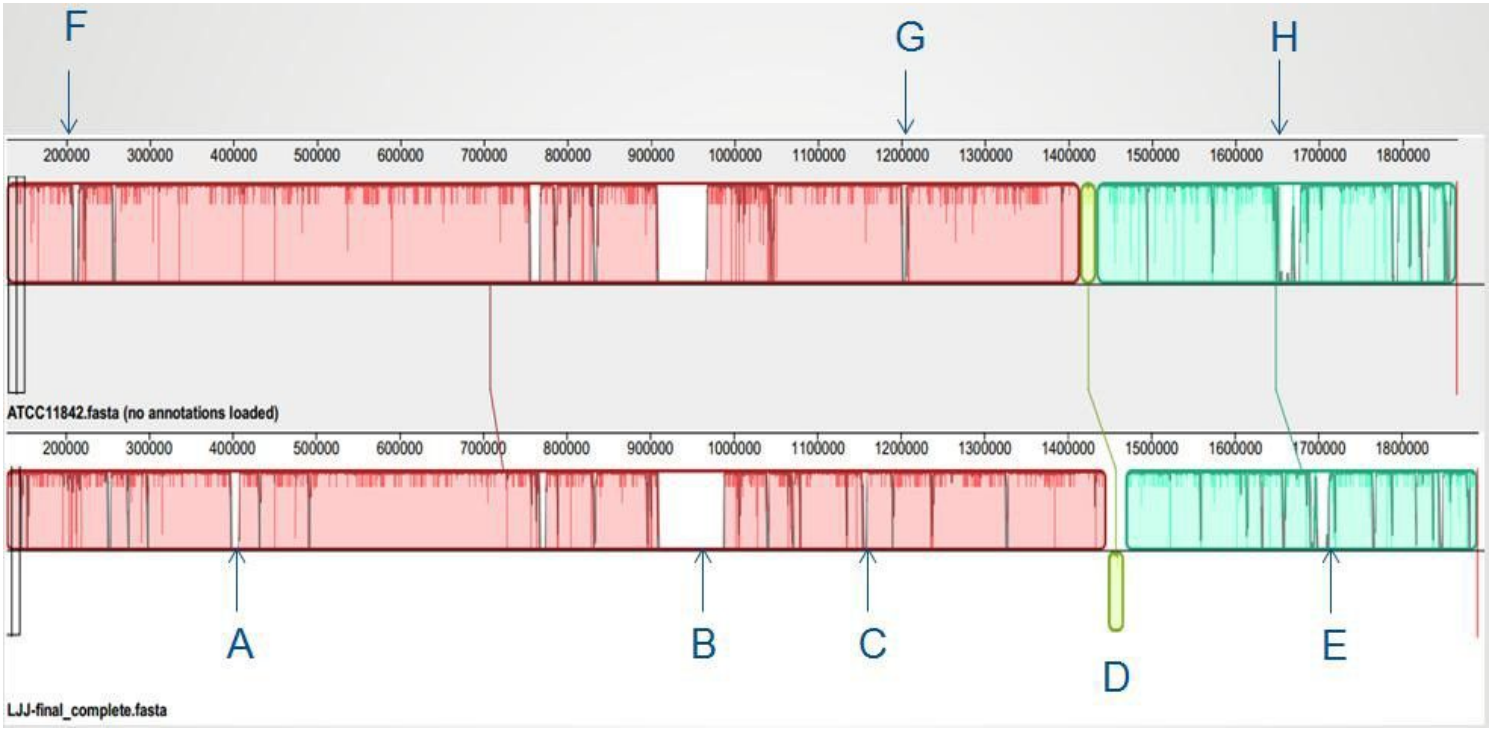


Figure 3

Colinearity analysis of LJJ and ATCC11842 genomes Compared with the type strain ATCC11842 genome, L. d. bulgaricus LJJ had an insertion or deletion sequence at ABCEFGH, and an inversion sequence at D.

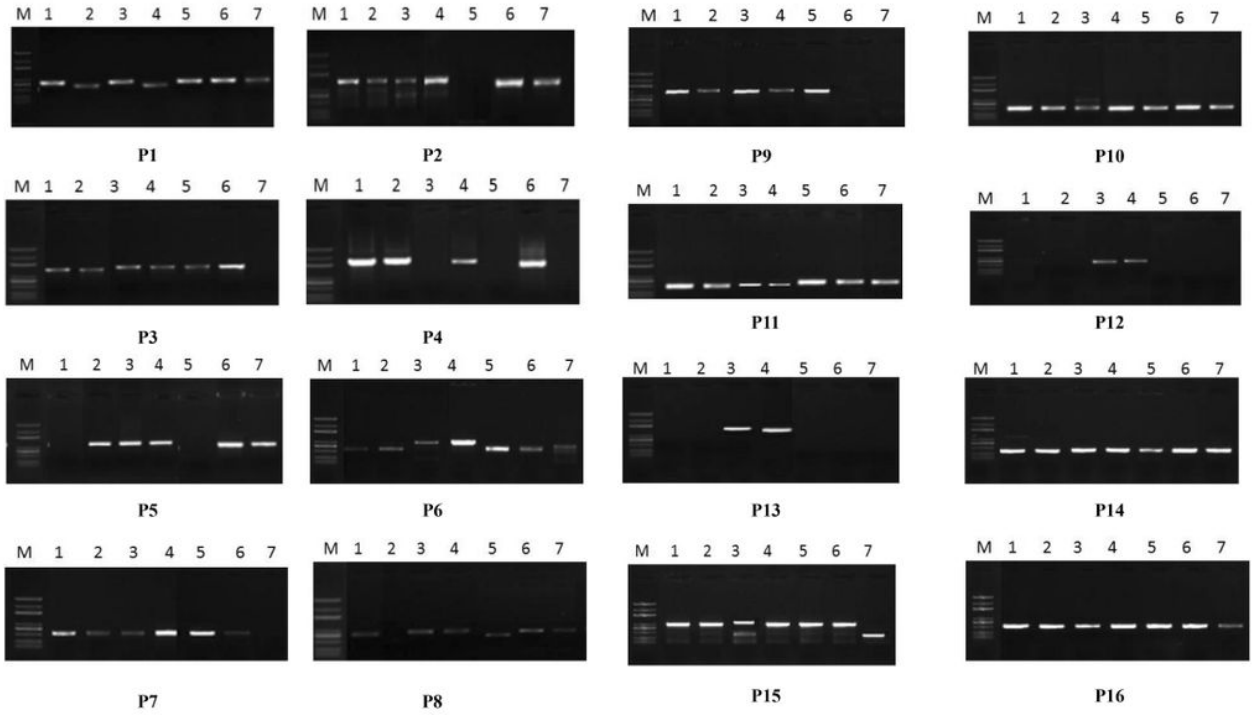
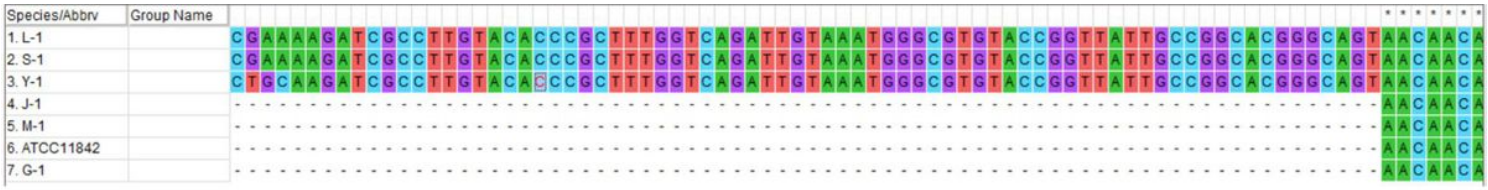
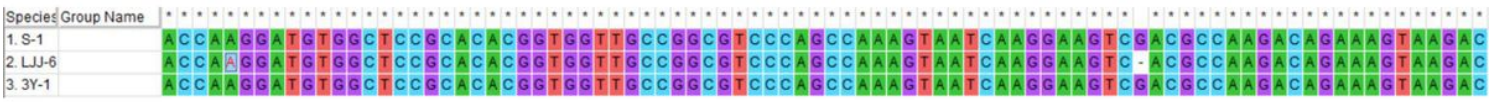


Figure 4

PCR amplification results of acid-tolerant genes from L. d. bulgaricus 1, 2, 3, 4, 5, 6, and 7 respectively indicate that 7 different L. d. bulgaricus are JB, GMC, SY3, LJJ, ATCC11842, M13, YL; M indicates that the marker strip is 100 bp, 250 bp, 500 bp, 750 bp, 1 k, 2 k, 3 k, 5 k from bottom to top. P1-P16 respectively indicate different primers referred to in Table 1.



a



b

c

Figure 5

Sequence analysis of acid-tolerant genes L means L. d. bulgaricus LJJ, S means L. d. bulgaricus SY3, Y means L. d. bulgaricus YL5, J means L. d. bulgaricus JB, M means L. d. bulgaricus M13 and G mean L. d. bulgaricus GMC, while the numbers represent sample numbers.

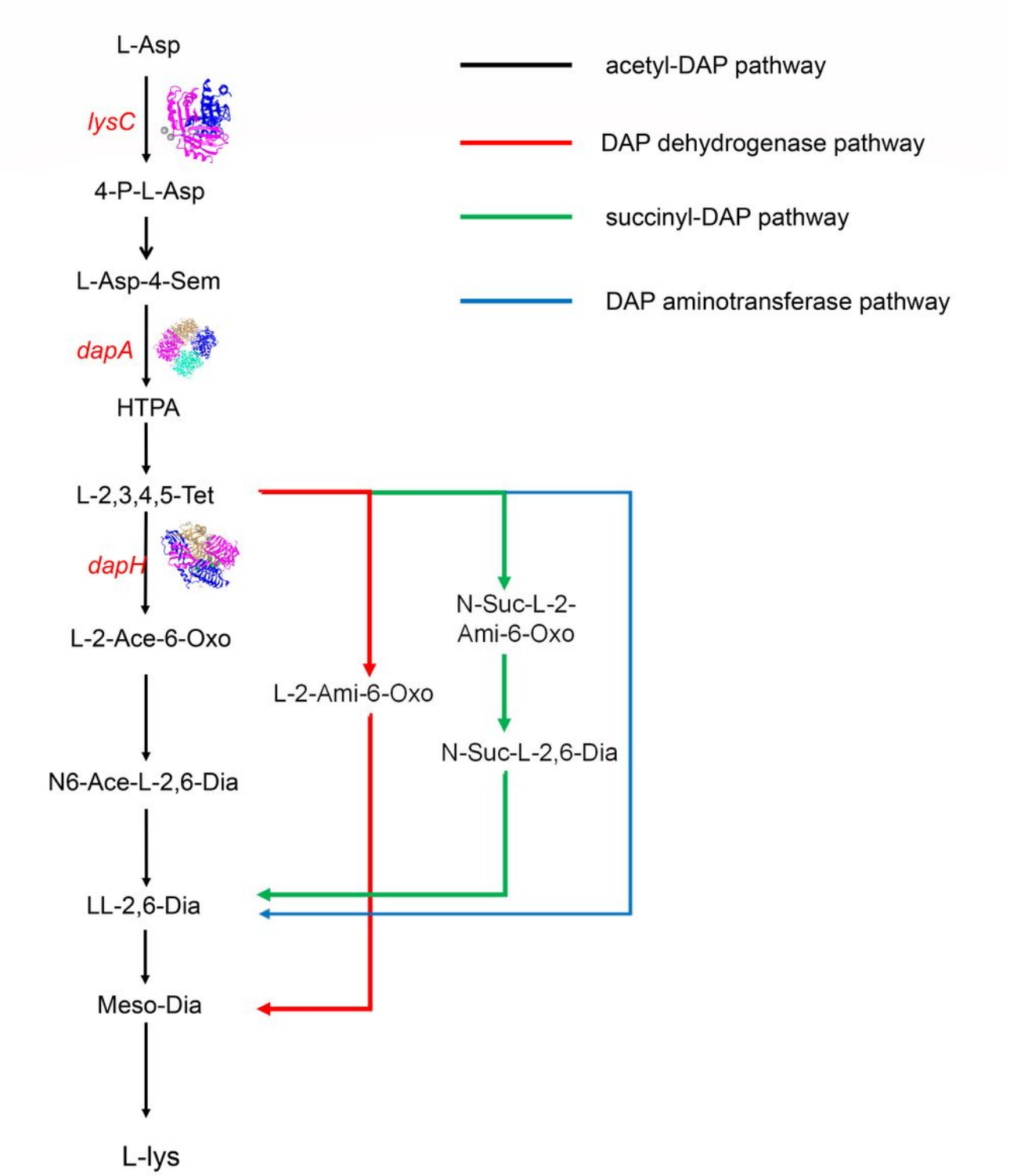


Figure 6

The regulation of Lysine synthesis The full name of the abbreviated name can be inquired in Abbreviations. lysC and dapA are the major key genes in the synthesis of lysine and are involved in all metabolic pathways, such as succinyl-DAP pathway, acetyl-DAP pathway, DAP dehydrogenase pathway, and DAP aminotransferase pathway. dapH encodes a tetrahydrodipicolinate N-acetyltransferase, a key enzyme in the acetyl-DAP pathway.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AdditionalFile2.fasta
- AdditionalFile4.ffn
- AdditionalFile7.pdf
- AdditionalFile3.faa
- AdditionalFile5.tif
- AdditionalFile1.tif
- AdditionalFile6.tif
- AdditionalFile8.tsv